Article

Improving Computer-Aided Cervical Cells Classification Using Transfer Learning Based Snapshot Ensemble

Wen Chen, Xinyu Li, Liang Gao and Weiming Shen *

State Key Lab. of Digital Manufacturing Equipment & Technology, Huazhong University of Science & Technology, Wuhan 430074, China; chenwen0704@hust.edu.cn (W.C.); lixinyu@hust.edu.cn (X.L.); gaoliang@hust.edu.cn (L.G.)

* Correspondence: shenwm@hust.edu.cn

Received: 28 September 2020; Accepted: 16 October 2020; Published: 19 October 2020

Abstract: Cervical cells classification is a crucial component of computer-aided cervical cancer detection. Fine-grained classification is of great clinical importance when guiding clinical decisions on the diagnoses and treatment, which remains very challenging. Recently, convolutional neural networks (CNN) provide a novel way to classify cervical cells by using automatically learned features. Although the ensemble of CNN models can increase model diversity and potentially boost the classification accuracy, it is a multi-step process, as several CNN models need to be trained respectively and then be selected for ensemble. On the other hand, due to the small training samples, the advantages of powerful CNN models may not be effectively leveraged. In order to address such a challenging issue, this paper proposes a transfer learning based snapshot ensemble (TLSE) method by integrating snapshot ensemble learning with transfer learning in a unified and coordinated way. Snapshot ensemble provides ensemble benefits within a single model training procedure, while transfer learning focuses on the small sample problem in cervical cells classification. Furthermore, a new training strategy is proposed for guaranteeing the combination. The TLSE method is evaluated on a pap-smear dataset called Herlev dataset and is proved to have some superiorities over the exiting methods. It demonstrates that TLSE can improve the accuracy in an ensemble manner with only one single training process for the small sample in fine-grained cervical cells classification.

Keywords: cervical cells classification; snapshot ensemble; transfer learning; deep convolutional neural network; deep learning; medical image

1. Introduction

Cervical cancer continues to be one of the prevalent cancers affecting women worldwide [1]. The disease is the most common cancer among women in 39 countries, and is the leading cause of cancer death in women in 45 countries [2]. The disease affects predominantly women in lower-resource countries, almost 70% of the global burden occurs in areas with low or medium levels of human development [2]. Manually screening abnormal cells from a pap-smear image has been a widely accepted method for prevention management as well as early detection of cervical cancer, especially in the developing countries [3]. However, manual assessment has many drawbacks in terms of being labor intensive, tedious, and time-consuming [3]. More importantly, the complex nature of cervical cell images presents significant challenges for manual screen analysis, the diagnosis results heavily rely on the experience of the technicians. Hence, the automation of cervical cells classification is essential for the development of a computer-aided classification system with low cost, adequate speed, and high accuracy [4], so that researchers and doctors can be released from the boring and repeated routine work. On the other hand, the computer-aided system can reduce the bias and provide robust results.

The computer-aided-systems were based on image analysis techniques during the past few decades, which follow a methodological framework with features derived from segmented cytoplasm and nuclei [4]. Based on the experience of skilled technicians and previous computer-aided-diagnosis, hand-crafted features [5,6], and engineered features [7–10] are taken as an vital component for the classification task. For example, a typical automated pap-smear analysis system consists of several steps: image acquisition, pre-processing, segmentation, feature extraction, feature selection, and classification [10]. A framework for automated detection and classification of cervical cancer from microscopic biopsy images using biologically interpretable features was proposed in [11]. Several feature selection methods are compared for diagnosis of cervical cancer by using a support vector machine classifier [12]. However, these hand-crafted features and engineered features have some limitations because they may not represent some complex discriminative information, and the feature selection process may have also neglected or excluded some potentially valuable information, which will eventually have some negative effects on the final results. Some other automated pap-smear analysis methods also prompted the development of automated cervical cancer screening, such as fuzzy C-means clustering technique [5], meta-heuristic algorithm [13], and Levenberg–Marquardt with adaptive momentum [14].

Recently, deep learning based methods have addressed many challenging problems in computer vision tasks effectively. Deep learning refers to methods that are designed to automatically learn and obtain high-dimensional features from raw data [15]. The superiority of convolutional neural networks has been demonstrated in plentiful image-processing tasks in the recent years, including object-detection [16,17], segmentation [18–21], human pose estimation [22], video classification [23], and image style transfer [24]. Furthermore, CNN models have achieved remarkable achievements in a broad range of diagnose tasks, including identifying moles from melanomas [25,26], referrals from fundus [27,28], optical coherence tomography images of the eye [29], breast lesion detection in mammograms [30] and breast cancer classification [31].

For cervical cells classification task, several approaches have been proposed towards the computer-aided systems and have already demonstrated some superior performance compared to the traditional methods. DeepPap was proposed to automatically extract deep features for classification [4]. A five-channel image dataset was formed, and these image patches were fed into the CNN models for classification [32]. A novel hybrid transfer learning technique was proposed to detect cervical cancer from cervix images [33]. A hierarchical modular neural network architecture, was reported for automated screening of cervical cancer [34]. These studies have greatly advanced the process in the field of cervical cells classification.

Despite of the recent significant progress in cervical cells classification, the high-precision accuracy remains a major obstacle for the fine-grained classification task. Most studies focus on the 2-class classification problem (normal or abnormal). A broad classification is useful for screening. Nonetheless, it is not enough for guiding clinical decisions on the diagnoses and treatments. Selection of treatment method depends on the phase of disease development. Fine-grained classification of cervical cells is of guiding significance to differentiate different disease phases and respective causes. Therefore, an appropriate treatment can be guaranteed. However, the CNN based approach achieved a highest accuracy of 98.3% [4] for the 2-class task on the Herlev dataset, while it only reached an accuracy of 64.1% for the 7-class task [32]. Thus, this indicates the need to find improved and efficient ways to classify cervical cells to finer categories automatically.

In our study, a transfer learning [35] based snapshot ensemble [36] method (TLSE) is proposed for the fine-grained cervical cells classification task. The snapshot ensemble aims at acquiring an ensemble result without a multi-step process, while the transfer learning provides a solution for the small sample cervical cells dataset with pre-trained model of good generalization. A new training strategy is conducted for the unification of the integration.

The original snapshot ensemble method provides an ensemble benefit within a single model training process instead of a traditional multi-step process, which lowers the computational cost.

This method has been used in some application areas, such as aerial scene classification [37] and fault diagnosis [38], but has not been tested or used for cervical cells classification. Furthermore, a disadvantage of snapshot ensemble is that the CNN model needs to be trained from scratch, which means that adequate amounts of raw samples are needed to ensure the training procedure. As for the cervical cells classification task, well-labeled data are limited because collecting high quality labeled data is very time consuming and costly. The small sample data may not provide enough information for training the whole model from scratch. On the other hand, transfer learning provides a solution to address this issue, which helps to eliminate the need of training the whole network all over again. The transfer learning method is widely adopted in several medical application areas, such as breast cancer diagnosis [39], epithelium–stroma classification [40], recognition of metastatic tissue in lymph node sections [41], lung cancer recognition [42], and cervical cells classification [4,32]. The proposed TLSE is meaningful to ensure the training of snapshot ensemble without the constraint of training the network with sufficient data from scratch. Furthermore, the advantages of the deep CNN model can also be leveraged into the cervical cells classification task, as the transfer learning phase makes the training easier.

The rest of the paper is organized as follows: a detailed description of the proposed transfer learning based snapshot ensemble (TLSE) is illustrated in Section 2. The experiment results conducted to evaluate the proposed method are reported in Section 3. The discussions and future work are presented in Section 4. The conclusion are summarized in Section 5.

2. Materials and Methods

2.1. Dataset and Pre-Processing

The Herlev dataset [6] was collected at Herlev University Hospital, the department of Pathology and department of Automation on Technical University of Denmark. These cervical cells images were collected and manually annotated into 7 classes by skilled cyto-technicians. The Herlev dataset is consist of 917 cell images, and each of the cell images only contains one single nuclei. The distribution of the Herlev pap-smear dataset is illustrated in Table 1 and some cell examples are provided in Figure 1.

Table 1. The distribution of Herlev dataset.

Cell Type		Numbers
Normal	Superficial squamous epithelial	74
	Intermediate squamous epithelial	70
	Columnar epithelial	98
Abnormal	Mild squamous non-keratinizing dysplasia	182
	Moderate squamous non-keratinizing dysplasia	146
	Severe squamous non-keratinizing dysplasia	197
	Squamous cell carcinoma in situ intermediate	150

The data pre-processing method presented in [4] is adopted in this paper for creating the training samples. To simplify the procedure, we only followed these steps: extract several image patches in size 128×128 by translating the centroid of the ground-truth nucleus mask. After that, in order to build a balanced training set, different rotation rates are set to abnormal cell images and normal ones, respectively. The rotation rate is set to 36 degree for individual abnormal cell images, and for each normal cell image, 18 degree is performed. These image patches are then resized to $224 \times 224 \times 3$ for facilitating the transfer learning phase. Zero padding is adopted to void regions that lie outside of the image boundary. The data augmentation step is crucial to the training of the deep CNN models,

which improves the accuracy and convergence when training deep CNN models. Figure 2 shows the two sets of cervical cell image patches by cropping and rotation.

Figure 1. The original cervical cell images on Herlev dataset.

Figure 2. Cervical cell images by cropping and rotation (left: abnormal; right: normal).

2.2. Transfer Learning Based Snapshot Ensemble Method (TLSE)

In this paper, a snapshot ensemble combined with transfer learning approach (TLSE) is proposed for the 7-classes cervical cells classification task. Different from the traditional ensemble methods that combined several CNN models together to get the ensemble benefits, the proposed method can obtain a comparable ensemble result within a single model training procedure. Furthermore, in order to address the small sample issue as well as explore the full capacity of deep CNN model, transfer learning is conducted ahead of snapshot ensemble. This is a new attempt to get ensemble benefits from deep convolutional neural network in medical application areas. This method will expand the application field of snapshot ensemble in cervical cells classification, as well as bring some new exploration on transfer learning by integrate it with an ensemble method. Furthermore, a new training strategy is proposed in order to guarantee the integration of these two methods. To sum up, the proposed approach integrates the transfer learning with snapshot ensemble based on deep CNN for the fine-grained cervical cells classification task.

This section describes the details of the transfer learning based snapshot ensemble (TLSE) method. An overview is reported in Figure 3.

Figure 3. Overview of the transfer learning based snapshot ensemble method (TLSE).

The transfer learning phase is conducted to fine-tune the pre-trained model towards the target dataset, while the snapshot ensemble phase is designed for getting several snapshot models within one single model training process. During the testing process, these saved snapshot models are combined together by averaging their soft-max outputs.

2.2.1. Transfer Learning Phase for TLSE

In our study, a CNN model is adapted as the base model, which is pre-trained on the ImageNet dataset (ILSVRC2012) [43]. The last fully connected layers of the downloaded model are excluded. Instead of using several data augmentation techniques to reduce the overfitting problem, the modification of model architecture is designed by adding regularization to the model. On top of the base model, three fully connected layers with 512, 128, and 7 nodes are conducted. The dropout [44] is adopted in the first two fully connected layers, which consists of setting the output of each hidden neuron to zero with a probability of 0.5. This technique forces CNN to learn robust features. The ReLu [45] activation is applied in the first two fully connected layers. The weights are initialized with he-style [46], which allows extremely deep CNN models to converge, especially when taking ReLu into account. Batch Normalization [47] is set before the activation layer to provide activation with the stable distribution, which can accelerate deep network training by reducing internal covariate shift. These modifications of model architecture allow the network to generalize to the new data distribution easily.

The training technique for the transfer learning phase in this TLSE method is that: firstly, freeze the pre-trained ImageNet's weight and remove the fully connected layers; secondly, add regularization to the model by making some specific refinements; at last, fine-tuning the whole CNN model using a small learning rate on the target dataset.

2.2.2. Snapshot Ensemble Phase for TLSE

Current practices indicate that combining the outputs of different models can achieve a better accuracy [48]. The outputs of several CNN models are combined by averaging their soft-max class posteriors. This improves the accuracy due to the complementarity and diversity of different CNN models. However, training multiple deep networks with different architecture or initial weights for model averaging is computationally expensive and time-consuming since it needs to train several CNNs, respectively. To remedy this time-wasting and extensive workload problem, snapshot ensemble was brought forward, which can obtain the similar goal of combining networks without additional training cost.

The snapshot ensemble phase for TLSE that is basically the same as the original snapshot ensemble. The snapshot ensemble develops an ensemble of accurate and diverse models within a single model training process. By using a cyclic learning rate schedule [49], snapshot ensemble enables the network to visit several local minima through the optimization process, and several model snapshots are saved at those local various minima.

The snapshot ensemble divides the whole training epochs into M cycles, and the learning rate is forced to decline at a very fast pace, for the sake of enabling the model to reach a local minimum. Thus, at every end of the cycles, a snapshot of the model weights is taken and saved. After that, the learning rate is raised for the start of the next training cycle. After M training cycles, M model snapshots are obtained, which will be used for the final ensemble. It is important to emphasize that the total training time is the same as a standard schedule. The ensemble prediction at the testing process is the average of the last two models' soft-max outputs.

2.2.3. The Training Strategy for TLSE

A novel training strategy is proposed for TLSE. Because there was a barrier in the combination of transfer learning and snapshot ensemble, these two methods cannot be combined directly. For the training strategy of transfer learning, the learning rate needs to be small for fine-tuning the model towards the target dataset. However, for the training strategy of snapshot ensemble, a big learning rate is required for helping the model to escape from the local minimum for the next cycle. This small-or-large learning rate issue makes model hard to train. In order to eliminate this obstacle, a new training strategy is conducted. Figure 4 illustrates the details.

Figure 4. The training strategy for TLSE.

The main differences between the original training strategy in snapshot ensemble and our proposed method are in the following aspects: Firstly, transfer learning is conducted in front of snapshot ensemble. A pre-trained model with several modifications is taken as the base model for snapshot ensemble. Secondly, two different learning rates are designed within one single optimization process, which are the initial learning rate and restart learning rate. The initial learning rate is been used for fine-tuning the pre-trained model towards the target dataset, while the restart learning rate is adopted to provoking the model and dislodge it from the minimum with the intension of restart training the model with a better initialization.

The initial learning rate is smaller than the restart learning rate. The smaller initial learning rate helps the pre-trained model adapted to the target dataset, while the large restart learning rate provides energy for the model to escape from a critical point. The two learning rates decline at a very fast pace with the strategy described in [49], which ensures the model to converge after only a few epochs. The very first cycle adopt the initial learning rate, while remain cycles use the restart learning rate. When the training of M cycles end, multiple convergences along with several well behaved local minimum at the end of every training cycle. At last, the outputs of the last two saved snapshot models are averaged as the final prediction.

2.2.4. Training and Testing Protocols

A pre-trained model has been downloaded and been used as a base model to train the neural network for cervical cells classification on the Herlev dataset. The proposed TLSE is conducted for fine-tuning the downloaded model towards the Herlev dataset as well as getting several snapshot models within the single model training process. During the testing process, the last two snapshots are integrated by averaging their soft-max outputs.

In the setting of the proposed approach, Stochastic Gradient Descent and Cross-entropy are used as the optimizer and loss function. The initial learning rate is set to 0.0001, and the restart learning rate is set to 0.2. The cyclic cosine annealing schedule is adopted to make the model converge to multiple local minima. The learning rate is updated at every iteration. The batch size is set to 32. The implementation is based on Keras platform, and the experiments are conducted on Ubuntu 16.04 with a single NVIDIA TITAN Xp GPU.

3. Results

In this study, several experiments were designed in order to show the effectiveness of the proposed approach.

3.1. The Classification Results of TLSE Method and Transfer Learning Method Based On Different CNN Models

The first experiment examined the impact of different CNN models on the TLSE method by comparing the classification results. Three different CNN architectures are chosen to be base models: VGG model [50], ResNet-18 [51], and Inception-ResNet [52] model. The architectures of these three pre-trained CNN models are shown in Figures 5–7, respectively. The aim of this experiment was to show the adaptability of the TLSE, whether it can be adopted in the training of different CNN models. The classification results of TLSE method based on these three CNN models are summarized in Table 2.

Table 2. The classification results of TLSE method and transfer learning method based on different models.

	VGG	ResNet	Inception-ResNet
Transfer learning method	62.24%	61.51%	64.50%
TLSE method	63.16%	62.69%	65.56%

Figure 5. The architecture of VGG model [50].

Figure 6. The architecture of ResNet model [51].

Figure 7. The architecture of Inception-ResNet model [52].

As shown in the Table 2, the result of Inception-ResNet model reaches the highest classification rate than other models. The VGG model and ResNet model get comparable results.

Secondly, to demonstrate the ensemble efficiency of the TLSE, a comparison experiment between TLSE method and the transfer learning method was designed.

The TLSE method utilizes transfer learning method and snapshot ensemble learning method together towards the Herlev dataset, while the transfer learning is adopted solely on the target dataset. The data augmentation techniques, model architecture refinements and the modification of last fully connected layers are the same for both TLSE based and transfer learning based methods. The data augmentation techniques taken in this experiment are crop and rotation. The VGG model, Inception-ResNet model and ResNet-18 model were adopted as the pre-trained models for both TLSE and transfer learning.

The comparison results of TLSE method and transfer learning method are reported in Table 2. By comparing the results with the TLSE method. There is a significant difference between the two groups, that TLSE based method achieves higher accuracies than the transfer learning method, suggesting that snapshot ensemble is crucial to achieve better representation. Table 2 also shows that the increments of accuracy are gained from all three models, which illustrates the robustness of the TLSE. The most striking result to merge from the data is that the TLSE method based on the Inception-ResNet model yield an accuracy of 65.56% on the pap-smear Herlev dataset, outperforming the single transfer learning method using the same base model.

3.2. The Classification Results of Different Architecture Refinements

The aim of this experiment was to analyze the effect of architecture refinements in the last two fully connected layers when the proposed TLSE method is adopted. We constructed some different architectures to show the effect of the dropout rate and batch normalization regularization of the CNN

model. We tested the classification rate with or without some model architecture refinements, which are the dropout layer and the batch normalization layer. Table 3 summarizes the comparison results.

Table 3. The classification results of different architecture refinements.

Model Architecture Refinement	Accuracy
With all the refinements	65.56%
Without dropout	63.16%
Without dropout and BN layer	62.98%

3.3. Comparison with other Methods

In order to show the effectiveness of TLSE, we selected three other previous deep learning based methods for comparison. There are two previous methods [6,32] reported the seven-class accuracy, three studies reported the overall error, and the methods in [4,32] are based on convolutional neural networks as well as transfer learning. The results are provided in Table 4. From the table, it can be seen that the TLSE achieves an accuracy of 65.56% and outperforms the other two methods (61.1% and 64.8%). The overall error of TLSE is higher than the DeepPap 1.6% [4], but lower than the benchmark 7.9% [6], the fine-grained CNN 7.7% [32], and Gen-wknn 3.9% [13]. These results suggest that the proposed TLSE method have some advantages in gaining better representation power, which results in a higher classification rate.

Table 4. Comparison of the proposed TLSE with other previous deep learning methods.

Method	Accuracy	Overall Error
DeepPap [4]	-(didn't report)	1.6% (5-fold validation)
Benchmark [6]	61.1%	7.9% (10-fold validation)
Gen-wknn [13]	-(didn't report)	3.9% (5-fold validation)
Fine-grained CNN [32]	64.8%	7.7% (5-fold validation)
TLSE	65.56%	3.6% (5-fold validation)

4. Discussion

The first set of analyses examined the robustness of the TLSE method on different CNN models by comparing the classification results. As shown in the Table 2, all three models get comparable results. It is evident from the table that the TLSE method works well with different CNN architectures on the Herlev dataset. The classification rate of Inception-ResNet model is higher than the other two models. The improvement of accuracy on the Inception-ResNet model is due to the availability of deeper CNN architecture, which helps the model to learn patterns that have more diversity and better representation power.

Table 2 also presents the transfer learning method with the same CNN models. By comparing the results with the TLSE method. There is a significant difference between these two methods, which is that the TLSE based method achieves higher accuracies. Turning now to the experimental evidence on proving that the addition of snapshot ensemble in TLSE has a positive influence on the fine-grained cervical cells classification task. Furthermore, these results also provide important insights into the robustness of the proposed TLSE method, as Table 2 illustrates that the increments of accuracies are gained from all three models. The most striking result to merge from the data is that the TLSE method based on the Inception-ResNet model yield an accuracy of 65.56% on the pap-smear Herlev dataset, outperforming the single transfer learning method using the same base model.

The positive influence of the TLSE to the final classification rate: get ensemble result through a single optimization process, add diversity and generalization to the CNN model.

The original snapshot ensemble is trained on a benchmark dataset cifar10, and several data augmentation techniques are adopted to increase the amount of data for providing sufficient information

for training deep CNN models. Nevertheless, in the medical image application area, the well-labelled data are not enough for training CNN model from scratch, and the training of CNN models will fall is a high probability event because of the over-fitting problem.

Fine-tuning the pre-trained CNN towards the target dataset can provide model with a strong ability of robustness and generalization. The pre-trained model with the fine-tuning phase can solve the problems of bad training approximation and pool generalization. The experiment results show that the TLSE method is effective and the established model offers a better accuracy and good capability of generalization.

The results obtained from Table 3 reveals that the model architecture refinements based on Inception-ResNet also improve classification accuracy. These minor refinements have a positive impact on the final classification accuracy. As expected, the combination of these techniques achieves a better accuracy compared to the non-modification architecture model, these architecture refinements help in capturing the high-level structural information, yielding better discriminability of different kinds of cervical cells. Further analyses reveal that the modification of the model architecture is effective and offers the model a better capability of generalization.

As reported in Table 4, the TLSE method reaches the highest accuracy on the Herlev datasets on the fine-grained classification task. On one hand, the snapshot ensemble phase in TLSE acquires an ensemble benefit by averaging the last two snapshot models' soft-max class posteriors. The ensemble result can be beneficial as it is a process of consulting several experts before taking a final decision. On the other hand, the transfer learning phase in TLSE provides better initialization weights rather than train from scratch, which provides the model with a good start point and makes the loss converge fast. In addition, the transfer learning plays a major role in reducing over-fitting, especially for small samples, as in the Herlev dataset case.

Despite of the accuracy improvement from TLSE, the proposed approach demonstrates a few limitations. Firstly, the size of input samples varies, so the images need to be resized into a fixed 128×128 for data processing, and zero padding is adopted for cropping. However, the cropped region may not contain the whole cell, and a fixed scale may not be appropriate when the cell image scales vary. Our ongoing study shows that this may be solved by using the spatial pyramid-pooling layer, for input images of arbitrary scales. Secondly, the final accuracy of the seven-class cervical cells classification is still not satisfied. The big gap between the two-class accuracy and the seven-class accuracy, as to our knowledge, is mainly due to the small sample. For the two-class problem, the amount of the data with data augmentation is sufficient. However, for the seven-class problem, the traditional data augmentation method may not be enough. As for the rare data especially in the medical image application areas, data augmentation methods, such as mix-up, GAN can be taken into consideration for increasing the number of samples. Thirdly, our approach is not based on microscopy slide images but rather on database images; thus, an important task related to image acquisition from the microscope is missing. The detection of cervical cells on microscopy slide images is another challenge task in this area. We anticipate that an end-to-end, highly accurate, and real-time cervical cells detection and classification system of this type is promising for the development of automation-assisted reading systems for cervical screening.

5. Conclusions

In this paper, a transfer learning based snapshot ensemble method called TLSE is proposed for fine-grained cervical cells classification task. The proposed TLSE approach integrates transfer learning with the snapshot ensemble in a coordinated way, a new training strategy has been proposed for this combination. Preponderances can be acquired from both snapshot ensemble and transfer learning. The proposed approach is able to gain better classification results in an ensemble manner using only one single training procedure, as well as address the small sample issue and prevent over-fitting during training the procedure. Since there is no need to train snapshot ensemble models from scratch, it can benefit the application areas of the snapshot ensemble method in image classification tasks,

especially when it comes to small samples. Furthermore, the representation power of the deep CNN model can be leveraged into the cervical cell classification task efficiently. Finally, some modifications are made to the architecture of the last two fully connected layers of the pre-trained model in order to add regularization to the CNN model. The TLSE method yields a comparably higher accuracy, compared to existing methods. However, more effort should be devoted to explore this fine-grained classification task for the clinical usage of automation-assisted reading systems for cervical screening.

Author Contributions: W.C. designed the algorithm, performed the experiments, and wrote the paper. X.L., L.G. and W.S. supervised the research. W.S. guided the paper writing. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Flores, B.E.; Acton, G.J. Older Hispanic women, health literacy, and cervical cancer screening. *Clin. Nurs. Res.* **2013**, *22*, 402–415. [\[CrossRef\]](#)
2. Stewart, B.W.; Wild, C.P. *World Cancer Report 2014*; International Agency for Research on Cancer: Lyon, France, 2014; p. 466.
3. Bengtsson, E.; Malm, P. Screening for Cervical Cancer Using Automated Analysis of PAP-Smears. *Comput. Math. Methods Med.* **2014**, *2014*, 842037. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Zhang, L.; Lu, L.; Nogues, I.; Summers, R.; Liu, S.; Yao, J. DeepPap: Deep Convolutional Networks for Cervical Cell Classification. *IEEE J. Biomed. Health Inform.* **2017**, *21*, 1633–1643. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Chankong, T.; Theera-Umpon, N.; Auephanwiriyakul, S. Automatic cervical cell segmentation and classification in Pap smears. *Comput. Methods Programs Biomed.* **2014**, *113*, 539–556. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Jantzen, J.; Norup, J.; Dounias, G.; Bjerregaard, B. Pap-smear Benchmark Data for Pattern Classification. In Proceedings of the NiSIS 2005: Nature Inspired Smart Information Systems (NiSIS), EU Co-Ordination Action, Albufeira, Portugal, 1–9 January 2005.
7. Nanni, L.; Lumini, A.; Brahnam, S. Local binary patterns variants as texture descriptors for medical image analysis. *Artif. Intell. Med.* **2010**, *49*, 117–125. [\[CrossRef\]](#)
8. Guo, Y.; Zhao, G.; Pietikäinen, M. Discriminative features for texture description. *Pattern Recognit.* **2012**, *45*, 3834–3843. [\[CrossRef\]](#)
9. Sokouti, B.; Haghipour, S.; Tabrizi, A.D. A framework for diagnosing cervical cancer disease based on feedforward MLP neural network and ThinPrep histopathological cell image features. *Neural Comput. Appl.* **2014**, *24*, 221–232. [\[CrossRef\]](#)
10. Win, K.P.; Kitjaidure, Y.; Hamamoto, K.; Aung, T.M. Computer-Assisted Screening for Cervical Cancer Using Digital Image Processing of Pap Smear Images. *Appl. Sci.* **2020**, *10*, 1800. [\[CrossRef\]](#)
11. Rajesh, K.; Rajeev, S.; Subodh, S. Detection and Classification of Cancer from Microscopic Biopsy Images Using Clinically Significant and Biologically Interpretable Features. *J. Med Eng.* **2015**, *2015*, 457906.
12. Ashok, B.; Aruna, P. Comparison of Feature selection methods for diagnosis of cervical cancer using SVM classifier. *Int. J. Eng. Res. Appl.* **2016**, *6*, 94–99.
13. Marinakis, Y.; Dounias, G.; Jantzen, J. Pap smear diagnosis using a hybrid intelligent scheme focusing on genetic algorithm based feature selection and nearest neighbor classification. *Comput. Biol. Med.* **2009**, *39*, 69–78. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Ampazis, N.; Dounias, G.; Jantzen, J. Pap-Smear Classification Using Efficient Second Order Neural Network Training Algorithms. In *Methods & Applications of Artificial Intelligence*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 230–245.
15. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [\[CrossRef\]](#) [\[PubMed\]](#)
16. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. *arXiv* **2017**, arXiv:1703.06870.
17. Ren, Y.; Zhu, C.; Xiao, S. Small Object Detection in Optical Remote Sensing Images via Modified Faster R-CNN. *Appl. Sci.* **2018**, *8*, 813. [\[CrossRef\]](#)
18. Chen, L.-C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking Atrous Convolution for Semantic Image Segmentation. *arXiv* **2017**, arXiv:1706.05587.

19. Tao, X.; Zhang, D.; Ma, W.; Liu, X.; Xu, D. Automatic Metallic Surface Defect Detection and Recognition with Convolutional Neural Networks. *Appl. Sci.* **2018**, *8*, 1575. [\[CrossRef\]](#)
20. Zhang, L.; Kong, H.; Liu, S.; Wang, T.; Chen, S.; Sonka, M. Graph-based segmentation of abnormal nuclei in cervical cytology. *Comput. Med Imaging Graph.* **2017**, *56*, 38–48. [\[CrossRef\]](#)
21. Panagiotakis, C.; Argyros, A. Region-based Fitting of Overlapping Ellipses and its application to cells segmentation. *Image Vis. Comput.* **2020**, *93*, 103810. [\[CrossRef\]](#)
22. Dang, Q.; Yin, J.; Wang, B.; Zheng, W. Deep learning based 2D human pose estimation: A survey. *Tsinghua Sci. Technol.* **2019**, *24*, 663–676. [\[CrossRef\]](#)
23. Karpathy, A.; Toderici, G.; Shetty, S.; Leung, T.; Sukthankar, R.; Fei-Fei, L. Large-Scale Video Classification with Convolutional Neural Networks. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1725–1732.
24. Gatys, L.A.; Ecker, A.S.; Bethge, M. Image Style Transfer Using Convolutional Neural Networks. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2414–2423.
25. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *542*, 115–118. [\[CrossRef\]](#)
26. Haenssle, H.; Fink, C.; Schneiderbauer, R.; Toberer, F.; Buhl, T.; Blum, A.; Kalloo, A.; Hassen, A.; Thomas, L.; Enk, A.; et al. Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann. Oncol.* **2018**, *29*, 1836–1842. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Gulshan, V.; Peng, L.; Coram, M.; Stumpe, M.C.; Wu, D.; Narayanaswamy, A.; Venugopalan, S.; Widner, K.; Madams, T.; Cuadros, J.; et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* **2016**, *316*, 2402–2410. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Poplin, R.; Varadarajan, A.V.; Blumer, K.; Liu, Y.; McConnell, M.V.; Corrado, G.S.; Peng, L.; Webster, D.R. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat. Biomed. Eng.* **2018**, *2*, 158–164. [\[CrossRef\]](#)
29. De Fauw, J.; Ledsam, J.R.; Romera-Paredes, B.; Nikolov, S.; Tomasev, N.; Blackwell, S.; Askham, H.; Glorot, X.; O'Donoghue, B.; Visentin, D.; et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat. Med.* **2018**, *24*, 1342–1350. [\[CrossRef\]](#)
30. Kooi, T.; Litjens, G.; Ginneken, B.; Gubern-Mérida, A.; Sánchez, C.; Mann, R.; Heeten, G.; Karssemeijer, N. Large Scale Deep Learning for Computer Aided Detection of Mammographic Lesions. *Med Image Anal.* **2016**, *35*, 303–312. [\[CrossRef\]](#)
31. Alkhaleefah, M.; Ma, S.-C.; Chang, Y.-L.; Huang, B.; Chittam, P.K.; Achhannagari, V.P. Double-Shot Transfer Learning for Breast Cancer Classification from X-Ray Images. *Appl. Sci.* **2020**, *10*, 3999. [\[CrossRef\]](#)
32. Lin, H.; Hu, Y.; Chen, S.; Yao, J.; Zhang, L. Fine-Grained Classification of Cervical Cells Using Morphological and Appearance Based Convolutional Neural Networks. *IEEE Access* **2019**, *7*, 71541–71549. [\[CrossRef\]](#)
33. Kudva, V.; Prasad, K.; Guruvare, S. Hybrid Transfer Learning for Classification of Uterine Cervix Images for Cervical Cancer Screening. *J. Digit. Imaging* **2019**, 1–13. [\[CrossRef\]](#)
34. Ali, M.; Sarwar, A.; Sharma, V.; Suri, J. Artificial neural network based screening of cervical cancer using a hierarchical modular neural network architecture (HMNNA) and novel benchmark uterine cervix cancer database. *Neural Comput. Appl.* **2019**, *31*, 2979–2993. [\[CrossRef\]](#)
35. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [\[CrossRef\]](#)
36. Huang, G.; Li, Y.; Pleiss, G.; Liu, Z.; Hopcroft, J.E.; Weinberger, K.Q. Snapshot Ensembles: Train 1, get M for free. *arXiv* **2017**, arXiv:1704.00109.
37. Dede, M.A.; Aptoula, E.; Genc, Y. Deep Network Ensembles for Aerial Scene Classification. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 732–735. [\[CrossRef\]](#)
38. Wen, L.; Gao, L.; Li, X. A New Snapshot Ensemble Convolutional Neural Network for Fault Diagnosis. *IEEE Access* **2019**, *7*, 32037–32047. [\[CrossRef\]](#)
39. Samala, R.K.; Chan, H.; Hadjiiski, L.; Helvie, M.A.; Richter, C.D.; Cha, K.H. Breast Cancer Diagnosis in Digital Breast Tomosynthesis: Effects of Training Sample Size on Multi-Stage Transfer Learning Using Deep Neural Nets. *IEEE Trans. Med. Imaging* **2019**, *38*, 686–696. [\[CrossRef\]](#) [\[PubMed\]](#)

40. Huang, Y.; Zheng, H.; Liu, C.; Ding, X.; Rohde, G.K. Epithelium-Stroma Classification via Convolutional Neural Networks and Unsupervised Domain Adaptation in Histopathological Images. *IEEE J. Biomed. Health Inform.* **2017**, *21*, 1625–1632. [[CrossRef](#)] [[PubMed](#)]
41. Kandel, I.; Castelli, M. How Deeply to Fine-Tune a Convolutional Neural Network: A Case Study Using a Histopathology Dataset. *Appl. Sci.* **2020**, *10*, 3359. [[CrossRef](#)]
42. Lin, C.-J.; Jeng, S.-Y.; Chen, M.-K. Using 2D CNN with Taguchi Parametric Optimization for Lung Cancer Recognition from CT Images. *Appl. Sci. Basel* **2020**, *10*, 2591. [[CrossRef](#)]
43. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 2012. [[CrossRef](#)]
44. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
45. Nair, V.; Hinton, G.E. Rectified Linear Units Improve Restricted Boltzmann Machines. In Proceedings of the 27th International Conference on Machine Learning (ICML-10), Haifa, Israel, 21–24 June 2010; pp. 807–814.
46. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *arXiv* **2015**, arXiv:1502.01852.
47. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv* **2015**, arXiv:1502.03167.
48. Caruana, R.; Niculescu-Mizil, A.; Crew, G.; Ksikes, A. Ensemble Selection from Libraries of Models. In Proceedings of the International Conference on Machine Learning, Banff, AB, Canada, 4–8 July 2004.
49. Loshchilov, I.; Hutter, F. SGDR: Stochastic Gradient Descent with Restarts. *arXiv* **2016**, arXiv:1608.03983.
50. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
51. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv* **2015**, arXiv:1512.03385.
52. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. *arXiv* **2016**, arXiv:1602.07261.

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).